

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS ✓
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF: Takao SHINOZAWA, et al.

GAU:

SERIAL NO: NEW APPLICATION

EXAMINER:

FILED: HERewith

FOR: HOMOLOGU ANALYSIS SYSTEM, HOMOLOGU ANALYSIS METHOD, HOMOLOGU ANALYSIS PROGRAM, AND TRANSACTION ESTABLISHMENT SYSTEM

REQUEST FOR PRIORITY

COMMISSIONER FOR PATENTS  
ALEXANDRIA, VIRGINIA 22313

SIR:

- ☒ Full benefit of the filing date of International Application Number PCT/JP01/10217, filed November 22, 2001, is claimed pursuant to the provisions of 35 U.S.C. §120.
- ☐ Full benefit of the filing date(s) of U.S. Provisional Application(s) is claimed pursuant to the provisions of 35 U.S.C. §119(e):  
Application No. Date Filed
- ☒ Applicants claim any right to priority from any earlier filed applications to which they may be entitled pursuant to the provisions of 35 U.S.C. §119, as noted below.

In the matter of the above-identified application for patent, notice is hereby given that the applicants claim as priority:

<u>COUNTRY</u>	<u>APPLICATION NUMBER</u>	<u>MONTH/DAY/YEAR</u>
Japan	2001-183856	June 18, 2001
Japan	2001-231810	July 31, 2001

Certified copies of the corresponding Convention Application(s)

- ☒ are submitted herewith
- ☐ will be submitted prior to payment of the Final Fee
- ☐ were filed in prior application Serial No. filed
- ☐ were submitted to the International Bureau in PCT Application Number  
Receipt of the certified copies by the International Bureau in a timely manner under PCT Rule 17.1(a) has been acknowledged as evidenced by the attached PCT/IB/304.
- ☐ (A) Application Serial No.(s) were filed in prior application Serial No. filed ; and
- ☐ (B) Application Serial No.(s)  
☐ are submitted herewith  
☐ will be submitted prior to payment of the Final Fee

Respectfully Submitted,

OBLON, SPIVAK, McCLELLAND,  
MAIER & NEUSTADT, P.C.

Gregory J. Maier

Registration No. 25,599

Robert T. Pous

Registration No. 29,099

Customer Number

22850

Tel. (703) 413-3000  
Fax. (703) 413-2220  
(OSMMN 05/03)

GJM:fb1

I:\USER\FBLAZ\PCT BY-PASS\246628.REQ.PRIORITY.DOC

日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日                      2001年 7月31日  
Date of Application:

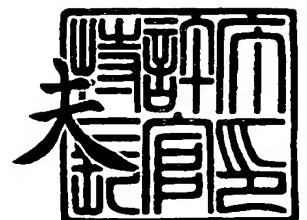
出願番号                      特願2001-231810  
Application Number:  
[ST. 10/C]:                      [JP 2001-231810]

出願人                      科学技術振興事業団  
Applicant(s):

2003年10月22日

特許庁長官  
Commissioner,  
Japan Patent Office

今井 康



【書類名】 特許願

【整理番号】 A000104178

【特記事項】 特許法第 3 0 条第 1 項の規定の適用を受けようとする特  
許出願

【提出日】 平成13年 7月31日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/00

【発明の名称】 相同性解析システム、相同性解析方法、相同性解析プロ  
グラム及び取引成立システム

【請求項の数】 15

【発明者】

    【住所又は居所】 埼玉県熊谷市久下 2 0 3 8 - 3

    【氏名】 篠沢 隆雄

【発明者】

    【住所又は居所】 群馬県桐生市東久方町 2 - 3 - 1 2 城西荘 7 号室

    【氏名】 堀池 徳祐

【発明者】

    【住所又は居所】 群馬県桐生市菱町 4 - 2 3 5 9 - 1 藤崎荘 8 号室

    【氏名】 濱田 一男

【発明者】

    【住所又は居所】 奈良県生駒市高山町 8 9 1 6 - 5

    【氏名】 金谷 重彦

【特許出願人】

    【識別番号】 396020800

    【氏名又は名称】 科学技術振興事業団

## 【代理人】

【識別番号】 100058479  
【弁理士】  
【氏名又は名称】 鈴江 武彦  
【電話番号】 03-3502-3181

## 【選任した代理人】

【識別番号】 100092196  
【弁理士】  
【氏名又は名称】 橋本 良郎

## 【選任した代理人】

【識別番号】 100091351  
【弁理士】  
【氏名又は名称】 河野 哲

## 【選任した代理人】

【識別番号】 100088683  
【弁理士】  
【氏名又は名称】 中村 誠

## 【先の出願に基づく優先権主張】

【出願番号】 特願2001-183856  
【出願日】 平成13年 6月18日

## 【手数料の表示】

【予納台帳番号】 011567  
【納付金額】 21,000円

## 【提出物件の目録】

【物件名】 明細書 1  
【物件名】 図面 1  
【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 相同性解析システム、相同性解析方法、相同性解析プログラム及び取引成立システム

【特許請求の範囲】

【請求項 1】 解析対象となる解析対象データ群が、前記解析対象データ群とは異なる第 1 のデータ群あるいは第 2 のデータ群のいずれに類似するかを解析するための相同性解析システムであって、

前記解析対象データ群と第 1 のデータ群のそれぞれのデータ群に含まれるデータの相同性を示す第 1 の相同値  $x$  を算出するものであって、相同性の有無の判定境界を示すしきい値  $E$  を  $n$  個設定して各しきい値  $E_i$  ( $i = 1, 2, \dots, n$ ) 毎に第 1 の相同値  $x_i$  を算出する第 1 の相同値算出手段と、

前記解析対象データ群と第 2 のデータ群のそれぞれのデータ群に含まれるデータの相同性を示す第 2 の相同値  $y$  を算出するものであって、相同性の有無の判定境界を示すしきい値  $E$  を  $n$  個設定して各しきい値  $E_i$  ( $i = 1, 2, \dots, n$ ) 毎に第 2 の相同値  $y_i$  を算出する第 2 の相同値算出手段と、

第 1 の相同値  $x_i$  及び第 2 の相同値  $y_i$  並びにしきい値の数  $n$  との関係に基づいて、前記解析対象データ群が第 1 のデータ群又は第 2 のデータ群のいずれに類似するかを判定する相同性決定手段と

を具備してなることを特徴とする相同性解析システム。

【請求項 2】 前記第 1 の相同値算出手段は、

前記解析対象データ群と第 1 のデータ群のそれぞれのデータ群に含まれるデータの相同性が所定のしきい値  $E$  以上である場合には相同性有りと判定し、相同性有りとなるデータの数を第 1 の相同値  $x$  として算出し、

前記第 2 の相同値算出手段は、

前記解析対象データ群と第 2 のデータ群のそれぞれのデータ群に含まれるデータの相同性が所定のしきい値  $E$  以上である場合には相同性有りと判定し、相同性有りとなるデータの数を第 1 の相同値  $y$  として算出する

ことを特徴とする請求項 1 に記載の相同性解析システム。

【請求項 3】 前記第 1 のデータ群は  $n_A$  個のデータを有し、前記第 2 のデ

ータ群は  $n_B$  個のデータを有する場合、

前記第 1 の相同値算出手段は、一つのしきい値  $E_i$  に対して前記第 1 のデータ群の各データ毎に第 1 の相同値  $x_{ij}$  ( $j = 1, 2, \dots, n_A$ ) を算出し、その算出されたしきい値  $E_i$  についての第 1 の相同値  $x_i$  の平均値  $\bar{x}_i$  を算出し、

前記第 2 の相同値算出手段は、一つのしきい値  $E_i$  に対して前記第 2 のデータ群の各データ毎に第 1 の相同値  $y_{ij}$  ( $j = 1, 2, \dots, n_B$ ) を算出し、その算出されたしきい値  $E_i$  についての第 2 の相同値  $y_i$  の平均値  $\bar{y}_i$  を算出し、

前記相同性決定手段は、

【数 1】

$$u_i = \sqrt{\frac{1}{n_A + n_B - 2} \left\{ \sum_{j=1}^{n_A} (x_{ij} - \bar{x}_i)^2 + \sum_{k=1}^{n_B} (y_{ik} - \bar{y}_i)^2 \right\}}$$

とした場合に、以下の等式

【数 2】

$$Z_i^{(1)} = \frac{\bar{x}_i - \bar{y}_i}{u_i} \cdot \sqrt{\frac{n_A \cdot n_B}{n_A + n_B}} \quad (i = 1, 2, \dots, n)$$

により、第 1 のデータ群あるいは第 2 のデータ群のいずれに類似するかを示す相同性決定値  $Z_i^{(1)}$  を算出することを特徴とする請求項 1 に記載の相同性解析システム。

【請求項 4】 前記相同性決定値  $Z_i^{(1)}$  が  $t$  分布に従い、自由度  $\alpha$  とすると、 $t_\alpha(0, 10)$  よりも大きい場合には前記解析対象データ群は前記第 1 のデータ群と相同性を持つデータ数が多いと判定する判定結果導出手段をさらに備えることを特徴とする請求項 3 に記載の相同性解析システム。

【請求項 5】 前記相同性決定値  $Z_i^{(1)}$  が  $t$  分布に従い、自由度  $\alpha$  とすると、 $-t_\alpha(0, 10)$  よりも小さい場合には前記解析対象データ群は前記第 2 のデータ群と相同性を持つデータ数が多いと判定する判定結果導出手段をさらに備えることを特徴とする請求項 2 に記載の相同性解析システム。

【請求項 6】 前記判定結果導出手段はさらに、 $s$  を  $Z_i^{(1)}$  の標準偏差、 $\bar{Z}_i^{(1)}$  を  $Z_i^{(1)}$  の平均値とすると、

【数 3】

$$Z^{(2)} = \frac{|\overline{Z^{(1)}}| - t_{n_A+n_B-2}(0.10)}{s/\sqrt{(n-1)}}$$

の等式で示される相同性可否決定値  $Z^{(2)}$  を算出し、該相同性可否決定値  $Z^{(2)}$  が所定の値  $t_{n-1}(0, 10)$  未満の場合には相同性決定値  $Z_i^{(1)}$  が有効な値でないと判定することを特徴とする相同性可否判定手段を備えることを特徴とする請求項 4 又は 5 に記載の相同性解析システム。

【請求項 7】 前記自由度  $\alpha$  は、 $n_A + n_B - 2$  であることを特徴とする請求項 4 乃至 6 に記載の相同性解析システム。

【請求項 8】 前記第 1 及び第 2 の相同値算出手段は、BLAST 法により相同値  $x_i$  及び  $y_i$  を算出することを特徴とする請求項 1 乃至 7 に記載の相同性解析システム。

【請求項 9】 前記解析対象データ群、第 1 のデータ群及び第 2 のデータ群は、遺伝子の配列を示すデータであることを特徴とする請求項 1 乃至 8 に記載の相同性解析システム。

【請求項 10】 解析対象となる解析対象データ群が、前記解析対象データ群とは異なる第 1 のデータ群あるいは第 2 のデータ群のいずれに類似するかを解析するための相同性解析方法であって、

前記解析対象データ群と第 1 のデータ群のそれぞれのデータ群に含まれるデータの相同性を示す第 1 の相同値  $x$  を算出するステップであって、この第 1 の相同値  $x$  を、相同性の有無の判定境界を示すしきい値  $E$  を  $n$  個設定して各しきい値  $E_i$  ( $i = 1, 2, \dots, n$ ) 毎に第 1 の相同値  $x_i$  として算出するステップと、

前記解析対象データ群と第 2 のデータ群のそれぞれのデータ群に含まれるデータの相同性を示す第 2 の相同値  $y$  を算出するステップであって、この第 2 の相同値  $y$  を、相同性の有無の判定境界を示すしきい値  $E$  を  $n$  個設定して各しきい値  $E_i$  ( $i = 1, 2, \dots, n$ ) 毎に第 2 の相同値  $y_i$  として算出するステップと、

第 1 の相同値  $x_i$  及び第 2 の相同値  $y_i$  並びにしきい値の数  $n$  との関係に基づいて、前記解析対象データ群が第 1 のデータ群又は第 2 のデータ群のいずれに類似



するかを判定するステップと

を有することを特徴とする相同性解析システム。

【請求項 11】 解析対象となる解析対象データ群が、前記解析対象データ群とは異なる第 1 のデータ群あるいは第 2 のデータ群のいずれに類似するかを解析するための相同性解析プログラムであって、コンピュータを、

前記解析対象データ群と第 1 のデータ群のそれぞれのデータ群に含まれるデータの相同性を示す第 1 の相同値  $x$  を算出するものであって、相同性の有無の判定境界を示すしきい値  $E$  を  $n$  個設定して各しきい値  $E_i$  ( $i = 1, 2, \dots, n$ ) 毎に第 1 の相同値  $x_i$  を算出する第 1 の相同値算出手段と、

前記解析対象データ群と第 2 のデータ群のそれぞれのデータ群に含まれるデータの相同性を示す第 2 の相同値  $y$  を算出するものであって、前記しきい値  $E$  を  $n$  個設定して各しきい値  $E_i$  ( $i = 1, 2, \dots, n$ ) 毎に第 2 の相同値  $y_i$  を算出する第 2 の相同値算出手段と、

第 1 の相同値  $x_i$  及び第 2 の相同値  $y_i$  並びにしきい値の数  $n$  との関係に基づいて、前記解析対象データ群が第 1 のデータ群又は第 2 のデータ群のいずれに類似するかを判定する相同性決定手段

として機能させることを特徴とする相同性解析プログラム。

【請求項 12】 第 1 の取引当事者と少なくとも 2 人の第 2 の取引当事者との間での取引の成立を判定すべく、第 1 の取引当事者の少なくとも 2 つの取引条件データからなる取引条件が、第 2 の取引当事者のいずれの提示した少なくとも 2 つの取引条件データからなる取引条件に類似するかを解析して取引の成立の判定を行う取引成立システムであって、

第 1 の取引当事者の取引条件データと第 2 の取引当事者のうちの一人の取引条件データの相同性を示す第 1 の相同値  $x$  を算出するものであって、相同性の有無の判定境界を示すしきい値  $E$  を  $n$  個設定して各しきい値  $E_i$  ( $i = 1, 2, \dots, n$ ) 毎に第 1 の相同値  $x_i$  を算出する第 1 の相同値算出手段と、

第 1 の取引当事者の少なくとも 2 つの取引条件データと第 1 の相同値算出手段の相同値算出の対象とされなかった他の一人の取引条件データの相同性を示す第 2 の相同値  $y$  を算出するものであって、相同性の有無の判定境界を示すしきい値

E を  $n$  個設定して各しきい値  $E_i$  ( $i = 1, 2, \dots, n$ ) 毎に第 2 の相同値  $y_i$  を算出する第 2 の相同値算出手段と

を具備してなり、第 1 の相同値  $x_i$  及び第 2 の相同値  $y_i$  に基づいて取引成立の可否を判定することを特徴とする取引成立システム。

【請求項 1 3】 第 1 の相同値  $x_i$  及び第 2 の相同値  $y_i$  並びにしきい値の数  $n$  との関係に基づいて、第 1 の取引当事者の取引条件がいずれの第 2 の取引当事者の取引条件に類似するかを判定する取引成立判定手段をさらに備えることを特徴とする請求項 1 2 に記載の取引成立システム。

【請求項 1 4】 第 1 の取引当事者の一の取引条件データは第 2 の取引当事者の一の取引条件データに対応付けられており、

第 1 の相同値算出手段及び第 2 の相同値算出手段は、対応付けられた取引条件データ同士について相同値を算出することを特徴とする請求項 1 2 又は 1 3 に記載の取引成立システム。

【請求項 1 5】 前記取引成立判定手段は、第 1 の相同値  $x_i$  及び第 2 の相同値  $y_i$  並びにしきい値の数  $n$  との関係に基づいて、第 1 の取引当事者の取引条件に類似する取引条件を提示した第 2 の取引当事者を順序付けて導出することを特徴とする請求項 1 3 に記載の取引成立システム。

#### 【発明の詳細な説明】

##### 【0 0 0 1】

#### 【発明の属する技術分野】

本発明は、複数のデータからなるデータ群同士の相同性を解析するための相同性解析システム、相同性解析方法、相同性解析プログラム、取引成立システム及び取引成立システムに関する。

##### 【0 0 0 2】

#### 【従来の技術】

遺伝子の類似性を比較する従来の方法として最も普及している方法は、「系統樹解析」である。この方法は、遺伝子の類似性を含めた相互関係を最も厳密に解析することができる。しかし、多数の遺伝子に関する解析を行う場合は、それに要する時間が長い点において大きな欠点を要する。また、生物の固体間の全体的

な類似性、すなわち多数の遺伝子のマスとしての相関関係などの解析には、採用した遺伝子が本当にその生物間の関係を代表しているものであるかという点で、適当な遺伝子を選んでいるとは限らない。遺伝子の相同性・類似性を比較する方法としては、上記の系統樹解析に加えて、B L A S T法やF A S T A法が代表的である。ここで、B L A S T法とは、遺伝子のデータバンクの中からある特定の遺伝子と類似性を有する遺伝子を選別するプログラムを用いて解析を行う方法である。

### 【 0 0 0 3 】

#### 【発明が解決しようとする課題】

上記したようなB L A S T法やF A S T A法では、遺伝子間の類似性の度合い、すなわちどれほど類似性が高いかに応じて、どのような遺伝子がどの程度の類似性があるかということを解析し、遺伝子を選別・特定することができる。しかし、これらの方法は類似性の度合いを設定する領域、すなわちしきい値を固定しているために、他の類似性の度合いにおいて、どのような遺伝子が存在するかなどの解析を行う点で不十分であった。

### 【 0 0 0 4 】

本発明は上記課題を解決するためになされたもので、その目的とするところは、多数のデータ同士を定量的にかつ正確に比較することが可能な相同性解析システム、相同性解析方法、相同性解析プログラム及び取引成立システムを提供することにある。

### 【 0 0 0 5 】

#### 【課題を解決するための手段】

この発明のある観点によれば、解析対象となる解析対象データ群が、前記解析対象データ群とは異なる第1のデータ群あるいは第2のデータ群のいずれに類似するかを解析するための相同性解析システムであって、前記解析対象データ群と第1のデータ群のそれぞれのデータ群に含まれるデータの相同性を示す第1の相同値  $x$  を算出するものであって、相同性の有無の判定境界を示すしきい値  $E$  を  $n$  個設定して各しきい値  $E_i$  ( $i = 1, 2, \dots, n$ ) 毎に第1の相同値  $x_i$  を算出する第1の相同値算出手段と、前記解析対象データ群と第2のデータ群のそれぞれの

データ群に含まれるデータの相同性を示す第 2 の相同値  $y$  を算出するものであって、相同性の有無の判定境界を示すしきい値  $E$  を  $n$  個設定して各しきい値  $E_i$  ( $i = 1, 2, \dots, n$ ) 毎に第 2 の相同値  $y_i$  を算出する第 2 の相同値算出手段と、第 1 の相同値  $x_i$  及び第 2 の相同値  $y_i$  並びにしきい値の数  $n$  との関係に基づいて、前記解析対象データ群が第 1 のデータ群又は第 2 のデータ群のいずれに類似するかを判定する相同性決定手段とを具備してなることを特徴とする相同性解析システムが提供される。

#### 【0006】

このような構成によれば、多数のデータからなるデータ群同士の相同性、すなわち類似性をまとめて評価できる。また、データとして遺伝子データを適用することにより、多数の遺伝子群の相同性をまとめて評価できる。よって、従来のように系統樹を描くために代表する遺伝子を選択する必要が無くなるため、遺伝子群としての特徴をより正確に理解することが可能となる。

#### 【0007】

例えば、従来は多数の遺伝子群間の関係を調べる方法では、その遺伝子群を代表する遺伝子を数個選び出し、その系統樹を描く方法であった。しかし、この方法では選び出した遺伝子が本当に代表として適当であるか判別することができず、選び出した遺伝子が適切か否かで結果が大きく異なっていた。これに対して、対象とするデータ群であるすべての遺伝子 (ORF) について相同性検索を行い、あるしきい値を達している遺伝子 (ORF) 数を数える BLAST 法や FASTA 法によれば、互いの関係を推測することができる。しかしながら、この方法でも、遺伝子群にいくつかの異なった由来の遺伝子群が混じっていた場合には、しきい値の基準を変えると結果がかなり異なってしまう特徴を有していた。そこで、本発明のように、いくつものしきい値で相同性検索を行い、それぞれのしきい値において相同性のある遺伝子 (ORF) の数をプロットし、遺伝子群間の関係を推測する方法を取ることで、しきい値の適切性を考慮することなく極めて安定した相同性評価結果を得ることができる。

#### 【0008】

現在、ヒトゲノム解明以後の大きな流れとして、病気の原因となる遺伝子の選

別特定が重要となってきた。候補となった遺伝子を実際のヒトDNA試料からPCR遺伝子増幅法により、DNA断片として作成し、大腸菌などの発現系でその遺伝子が合成するタンパク質を実際に作らせ、その活性をみることや、あるいはそのタンパク質に対する抗体を作成することが行われる。抗体は実際にヒトの病理切片などで試すことで病気の検出方法などに有用である。また、病気の原因となる遺伝子の特定は患者の組織の病域の特定、あるいは健常人の遺伝子を解析することで、その病気にかかりやすい人を選別することができる。そうすることで、患者に対する治療の方法、例えば、保因者とそうでない人に対しては、例えばポリープ切除の緊急性の判断に応用できる。健常人に関しては、その病気に関する診断検査のそれぞれの人の遺伝子情報に基づき使い分けることが可能である。

#### 【0009】

また、本発明の別の観点によれば、第1の取引当事者と少なくとも2人の第2の取引当事者との間での取引の成立を判定すべく、第1の取引当事者の少なくとも2つの取引条件データからなる取引条件が、第2の取引当事者のいずれの提示した少なくとも2つの取引条件データからなる取引条件に類似するかを解析して取引の成立の判定を行う取引成立システムであって、第1の取引当事者の取引条件データと第2の取引当事者のうちの一人の取引条件データの相同性を示す第1の相同値 $x$ を算出するものであって、相同性の有無の判定境界を示すしきい値 $E$ を $n$ 個設定して各しきい値 $E_i$  ( $i=1, 2, \dots, n$ ) 毎に第1の相同値 $x_i$ を算出する第1の相同値算出手段と、第1の取引当事者の少なくとも2つの取引条件データと第1の相同値算出手段の相同値算出の対象とされなかった他の一人の取引条件データの相同性を示す第2の相同値 $y$ を算出するものであって、相同性の有無の判定境界を示すしきい値 $E$ を $n$ 個設定して各しきい値 $E_i$  ( $i=1, 2, \dots, n$ ) 毎に第2の相同値 $y_i$ を算出する第2の相同値算出手段とを具備してなり、第1の相同値 $x_i$ 及び第2の相同値 $y_i$ に基づいて取引成立の可否を判定することを特徴とする取引成立システムが提供される。

#### 【0010】

また、システム（装置）に係る本発明は、その装置により実現される方法の発

明としても成立する。

#### 【0 0 1 1】

また、装置または方法に係る本発明は、コンピュータに当該発明に相当する手順を実行させるための（あるいはコンピュータを当該発明に相当する手段として機能させるための、あるいはコンピュータに当該発明に相当する機能を実現させるための）プログラム、このプログラムを記録したコンピュータ読取り可能な記録媒体としても成立する。

#### 【0 0 1 2】

##### 【発明の実施の形態】

本発明の骨子は、解析対象となるデータ群（解析対象データ群）と、他のデータ群との相同性を算出し、解析対象データ群がいずれのデータ群と相同性を有するか否かを判定する点にある。

#### 【0 0 1 3】

以下、図面を参照しながら本発明の一実施形態を説明する。

#### 【0 0 1 4】

図 1 は本発明の第 1 実施形態に係る遺伝子解析システム 1 0 の全体構成を示す図である。図 1 に示すように、プロセッサ 1 と、このプロセッサ 1 に接続された出力手段 2 と、プロセッサ 1 に接続されたアミノ酸配列記憶手段 3 と、相同値記憶手段 4 から構成される。プロセッサ 1 は、B L A S T 解析手段 1 1 と T 検定手段 1 2 を備える。

#### 【0 0 1 5】

プロセッサ 1 が図示しない記録媒体から所定のプログラムを読み出すことにより各手段 1 1 及び 1 2 としてプロセッサ 1 が機能する。B L A S T 解析手段 1 1 は、複数のデータからなるあるデータ群と他のデータ群の相同性を相同値として算出する相同値算出手段を備える。また、T 検定手段 1 2 は、相同性決定値算出手段、相同性可否決定値算出手段及び判定結果導出手段を備える。相同性決定値算出手段は、相同性、すなわちあるデータ群と他のデータ群の類似性を数値により示した相同性決定値  $Z_i^{(1)}$  を算出する。相同性可否決定値算出手段は、相同性決定値算出手段で算出された相同性決定値  $Z_i^{(1)}$  が相同性を示す値とし

て適切であるか否かを判定するための数値である相同性可否決定値  $Z(2)$  を算出する。判定結果導出手段は、相同性決定値算出手段で算出された相同性決定値  $Z_i(1)$  と相同性可否決定値算出手段で算出された相同性可否決定値  $Z(2)$  に基づいていずれのデータ群に相同性の多いデータを有するかの判定を行う。

#### 【0016】

次に、図2を用いて本実施形態に係る遺伝子解析に必要な遺伝子データを取得するプロセスを説明する。

#### 【0017】

まず、インターネットなどのネットワークを介して、遺伝子解析システム10のBLAST解析手段11が酵母ORF (Open Reading Frames) アミノ酸配列を取得する(s21)。次に、この取得したアミノ酸配列を機能別43のカテゴリに分類する(s22)。そして、得られた機能別のカテゴリから、ミトコンドリアからの遺伝子移行の影響を受けていない酵母ORFを抽出する(s23)。なお、原核生物として、細胞寄生性の原核生物を除く原核生物のORFを15種類取得する。

#### 【0018】

以上により、真核生物である酵母のORFと、原核生物の15種類のORFが得られる。得られた各ORFは、アミノ酸配列記憶手段3に格納される。なお、これら各ORF (遺伝子データ) は、例えば酵母であれば酵母全体の各ORFがひとまとまりとして遺伝子データ群として格納され、例えば原核生物であれば、ある原核生物全体の各ORF (遺伝子データ) がひとまとまりとして遺伝子データ群として格納される。また、各ORFはアミノ酸配列により特定される。このアミノ酸配列記憶手段3に格納された遺伝子データ群の概念図を図3に示す。図3では、酵母ORFは全部で  $p$  個、古細菌AのORFは  $n_1$  個、古細菌BのORFは  $n_2$  個、真正細菌CのORFは  $n_3$  個、真正細菌DのORFは  $n_4$  個示してある。

#### 【0019】

次に、BLAST解析手段11は、以上のようにして得られた16種類のORFのうち、解析対象として酵母のORFを選択し、この解析対象である酵母ORF

Fと15種類のORFのそれぞれとの相同値ヒット数（オルソログス遺伝子数）をBLAST法を用いて算出する。

#### 【0020】

オルソログス遺伝子数を検出するには、まず酵母の全ORFと各細菌の全ORFに対してORF間の相同性の指標であるE-値（相同性と負の相関）を算出する。酵母と各細菌で互いに最小のE-値を持つORFをオルソログス遺伝子と判定する。オルソログス遺伝子の数（ヒット数）が生物間の相同性を表す。そして、様々なしきい値（E-値）に対してこのヒット数を算出する。

#### 【0021】

図4はBLAST法の概念図である。図4に示すように、酵母のORFを細胞寄生性の原核生物を除く15種類の原核生物のORFに対して相同性検索を行う。相同性検索は、まずプロセッサ1がアミノ酸配列記憶手段3から酵母のORF1と古細菌A<sub>1</sub>のORFを読み出す。そして、相同性の指標であるE-値を算出し、最小のE値を持つORFを古細菌A<sub>1</sub>の酵母ORF1との第一ヒットORFとし、そのORF名とその時のE値を記憶させる。これを全ORFに対して繰り返す。他の古細菌A<sub>2</sub>, ..., A<sub>n</sub>、真正細菌に対しても同様に行う。BLAST法については、例えば2001年2月1日発行の“Nature Cell Biology, Vol.3, No.2, pp210-214”に開示されている。また、遺伝子を比較する方法はBLAST法以外でも、アミノ酸配列同士、塩基配列同士、分子レベルなど、遺伝子を情報として特定できるいかなるレベルで比較してもよい。

#### 【0022】

さらに、図5はオルソログス遺伝子数で表される生物間の相同性の判定の概念を示す図である。図5では、矢印の先は第一ヒット遺伝子である。但し、互いのE値はしきい値を超えている。丸印で示したのは、互いにオルソログス遺伝子であることを示している。一方、酵母ORF2にとっての第一ヒット遺伝子は古細菌A<sub>1</sub>のORF3であり、古細菌A<sub>1</sub>のORF3にとっての第一ヒット遺伝子は酵母ORF3である。従って、酵母ORF2の第一ヒット遺伝子と古細菌A<sub>1</sub>のORF3の第一ヒット遺伝子は一致しない。オルソログス遺伝子の数が酵母と古細菌A<sub>1</sub>との生物間の相同性を表す。



## 【0023】

図5のように、酵母と古細菌A<sub>1</sub>に着目すると、この酵母のORF1の古細菌A<sub>1</sub>との第一ヒットORF-aとそのときのE-値を読み出し、そして古細菌A<sub>1</sub>のORF-aの酵母との第一ヒットORF- $\alpha$ とそのときのE-値を読み出す。ORF1とORF- $\alpha$ が一致し、両E-値がしきい値Eより大きいとき、オルソログ遺伝子と判定し、生物間相同値ヒット数としてカウントする。次に、酵母のORF2の古細菌A<sub>1</sub>との第一ヒットORF-bとそのときのE-値を読み出し、同様に古細菌A<sub>1</sub>とのオルソログ遺伝子の有無を判定する。ORF-bがオルソログ遺伝子のとき、生物間相同値ヒット数としてカウントする。

## 【0024】

このようにして、酵母ORF3～酵母ORF<sub>p</sub>についても同様にオルソログORFの有無を判定する。そして、得られたカウント値を古細菌A<sub>1</sub>との相同値ヒット数xとして相同値記憶手段4に格納する。

## 【0025】

以上は、酵母と古細菌A<sub>1</sub>に着目して説明したが、同様に酵母と古細菌A<sub>2</sub>, ..., A<sub>n</sub>についても行い、それぞれを酵母との相同値ヒット数xとしてそれぞれ相同値記憶手段4に格納する。

## 【0026】

また、プロセッサ1のBLAST解析手段11は、古細菌A<sub>1</sub>, ..., A<sub>n</sub>についての相同値ヒット数の平均値 $\bar{x}$ を算出し、得られた平均値を相同値記憶手段4に格納する。

## 【0027】

また、このような相同値ヒット数の算出を他の古細菌遺伝子データ群や、真正細菌遺伝子データ群などについても行う。

## 【0028】

また、p個の遺伝子数である酵母遺伝子データ群とn<sub>A</sub>個の古細菌の遺伝子データ群とのオルソログ遺伝子検出において、あるしきい値E<sub>i</sub>について、j番目の古細菌について得られた相同値ヒット数をx<sub>ij</sub>とする。

## 【0029】

また、 $p$  個の遺伝子数である酵母遺伝子データ群と  $n_B$  個の真正細菌の遺伝子データ群とのオルソログス遺伝子検出において、あるしきい値  $E_i$  について、 $k$  番目の真正細菌について得られた相同値ヒット数を  $y_{ik}$  とする。

#### 【0 0 3 0】

以上のようにして得られた相同値ヒット数としきい値  $E$  との関係を示したのが図 6 である。横軸は  $-\log E$ 、縦軸は相同値ヒット数である。なお、この図 6 では、それぞれの折れ線グラフが酵母遺伝子と比較する各細菌に対応している。類似性の度合いを厳しく、すなわちしきい値  $E$  を小さくしていくにつれ、条件を満たす遺伝子数、すなわち相同値ヒット数は漸減しているのが分かる。

#### 【0 0 3 1】

以上により取得された各種相同値ヒット数に基づき、プロセッサ 1 の T 検定手段は、以下による T 検定処理を行う。

#### 【0 0 3 2】

T 検定処理は、相同性決定値  $Z_i^{(1)}$  の算出、相同性可否決定値  $Z^{(2)}$  の算出及び判定結果導出の 3 つの処理からなる。

#### 【0 0 3 3】

まず、相同性決定値  $Z_i^{(1)}$  の算出手法について酵母遺伝子データ群が古細菌遺伝子データ群と真正細菌遺伝子データ群のいずれに相同性が高い遺伝子を多く共有するかを解析する場合について説明する。

#### 【0 0 3 4】

一例として、上述した各しきい値  $E_i$  ( $i = 1, 2, \dots, n$ ) において、有意水準 5 % の片側 T 検定をヒット数が少なくとも 5 以上という基準を満たす領域で行う。

#### 【0 0 3 5】

そして、相同性決定値  $Z_i^{(1)}$  を以下の等式により算出する。

#### 【0 0 3 6】

#### 【数 4】

$$Z_i^{(1)} = \frac{\bar{x}_i - \bar{y}_i}{u_i} \cdot \sqrt{\frac{n_A \cdot n_B}{n_A + n_B}} \quad (i = 1, 2, \dots, n)$$

## 【0 0 3 7】

なお、以上に示す等式で、 $x_{ij}$  は  $i$  番目の  $E$  値でのヒット数  $x_{ij}$  ( $j = 1, 2, \dots, n_A$ ) の平均値を、 $y_{ik}$  は  $i$  番目の  $E$  値でのヒット数  $y_{ik}$  ( $k = 1, 2, \dots, n_B$ ) の平均値を示す。 $n$  は、BLAST法で解析した際のしきい値の数  
を示している。また、不偏分散  $u_i$  は以下の等式で示される。

## 【0 0 3 8】

## 【数 5】

$$u_i = \sqrt{\frac{1}{n_A + n_B - 2} \left\{ \sum_{j=1}^{n_A} (x_{ij} - \bar{x}_i)^2 + \sum_{k=1}^{n_B} (y_{ik} - \bar{y}_i)^2 \right\}}$$

## 【0 0 3 9】

以上のようにして得られた相同性決定値  $Z_i(1)$  を統計処理したのが図 7 である。図 7 では、横軸に  $-\log E$  を、縦軸に相同性決定値  $Z_i(1)$  をとってあり、図 6 に示されるデータに基づいて得られた値である。相同性有り  
と判断された遺伝子数を 5 以上の領域で棒グラフにより示している。相同性決定値  $Z_i(1)$  が  $t_{n_A+n_B-2}(0, 10)$  ( $= 1.771$ ) 以上又は  $-t_{n_A+n_B-2}(0, 10)$  ( $= -1.771$ ) 以下の場合は有意水準 5 % (2 つの母平均が等しいと仮定したとき、標本平均が等しくない確率が 5 % 以下) で、各々のしきい値  $E$  に対して古細菌あるいは真正細菌の遺伝子群が出芽酵母の遺伝子群と類似性が高い。

## 【0 0 4 0】

次に、相同性可否決定値  $Z(2)$  の算出手法について説明する。この相同性可否決定値  $Z(2)$  の算出は、1 回目の  $T$  検定、すなわち相同性決定値  $Z_i(1)$  の算出における相同性決定値  $Z_i(1)$  が、 $t_{n_A+n_B-2}(0, 10)$  ( $= 1.771$ ) より設定した  $E_i$  全体に対して優位に大きいか、 $-t_{n_A+n_B-2}(0, 10)$  ( $= -1.771$ ) より小さいか、あるいはどちらの傾向も無いかを判定するために行われる。そのため、自由度  $n - 1$  において以下の等式により、相同性可否決定値  $Z(2)$  と  $t_{n-1}(0, 10)$  を計算する。

## 【0 0 4 1】

## 【数 6】

$$Z^{(2)} = \frac{|\overline{Z^{(1)}}| - t_{n_A+n_B-2}(0.10)}{s/\sqrt{(n-1)}}$$

## 【0042】

ここで、 $s$  は相同性決定値  $Z_i^{(1)}$  の標準偏差、 $Z_i^{(1)}$  は  $Z_i^{(1)}$  の平均値である。相同性可否決定値  $Z^{(2)}$  が基準値  $t_{n-1}$  より大きいときは  $Z_i^{(1)}$  が  $t_{n_A+n_B-2}$  よりも有意水準 5 % で大きいと判定できる。

## 【0043】

次に、以上により算出された相同性決定値  $Z_i^{(1)}$  及び相同性可否決定値  $Z^{(2)}$  に基づいて、判定結果導出を行う。判定結果導出は、図 8 に示す判定テーブルを用いて行う。図 8 に示すように、 $Z^{(2)} / t_{n-1}(0, 10)$  が 1 以上で、 $Z_i^{(1)}$  の平均値が正であれば、そのカテゴリーにおける酵母 ORF 群は真正細菌よりも古細菌との方に多くの相同性を有する ORF を含むと判定できる。また、 $Z^{(2)} / t_{n-1}(0, 10)$  が 1 以上で、 $Z_i^{(1)}$  の平均値が負であれば、そのカテゴリーにおける酵母の ORF 群は古細菌よりも真正細菌との方に多くの相同性を有する ORF を含むと判定できる。また、 $Z^{(2)} / t_{n-1}(0, 10)$  が 1 より小さい場合はいずれの細菌に相同性を有するかを判定できないと判定する。

## 【0044】

このような構成によれば、多数の遺伝子群の相同性をまとめて評価できる。よって、従来のように系統樹を描くために代表する遺伝子を選択する必要がなくなるため、遺伝子群としての特徴をより正確に理解することが可能となる。

## 【0045】

例えば、この方法を用いて真核生物の起源を探ることができる。従来は、主に rRNA と小数のタンパク質について系統樹を作成し、推測していた。そして、その推測によれば、DNA 複製、転写、翻訳等の遺伝子群は古細菌起源で、エネルギー代謝の遺伝子群はミトコンドリアの共生が起源で、その他の遺伝子群は古細菌と真正細菌のモザイク構造になっていると予想されていた。また、その他説

が分かれていた。

#### 【 0 0 4 6 】

そこで、酵母の持つそれぞれの機能の由来を考察するため、細胞寄生性細菌を除く 1 5 種のバクテリアの O R F (Open Reading Frames) を、機能別に分類された酵母の O R F と比較する。すなわち、上述した B L A S T 解析、T 検定の解析対象を酵母遺伝子データ群、比較対照を古細菌遺伝子データ群及び真正細菌遺伝子データ群とする。その結果、4 3 の機能別に分類された酵母の遺伝子群のうち、2 0 の遺伝子群は古細菌、又は真正細菌のどちらか一方に相同性の高い遺伝子を多く含むことが分かった。一般に、進化系統の近い生物間では相同性の高い遺伝子を多く保持していることが知られていることから、酵母遺伝子の機能カテゴリー別にそれぞれの起源を推定できる。

#### 【 0 0 4 7 】

具体的にこの B L A S T 解析及び T 検定を行うことにより、DNA 合成及び複製、転写、翻訳、減数分裂、細胞周期調節、小胞体形成、核形成などの遺伝子群は古細菌 O R F と、エネルギー代謝、各種物質代謝、細胞内への物質輸送、ストレス応答、解毒、イオンホメオスタシスの遺伝子群は真正細菌 O R F と、それぞれ相同性が高い遺伝子を多く共有することが分かった。

#### 【 0 0 4 8 】

従って、核（遺伝情報）に関連した遺伝子は古細菌の遺伝子が、また細胞質（ホメオスタシス）に関連した遺伝子は真正細菌の遺伝子が起源であると考えられる。また、このことは、どちらかの遺伝子群が後から他方の生物に置き換わったことを示す。

#### 【 0 0 4 9 】

これらの結果は、真核生物の核は古細菌の真正細菌への共生に由来することを示唆する。すなわち、相同性判定により遺伝子群間の関係を推測することにより、真核生物は真正細菌へ古細菌が共生することによって誕生したと考えられる可能性が高い。

#### 【 0 0 5 0 】

このことを現在有力とされる各説に応じて考えると、以下のようになる。

## 【 0 0 5 1 】

まず、真正細菌由来の遺伝子はミトコンドリア共生の結果もたらされたのではないかとする説に対しては、ミトコンドリア関連遺伝子は解析のデータから削除されており、影響しないと考えられると結論づけられる。

## 【 0 0 5 2 】

また、遺伝子水平伝達の蓄積ではないかとする説に対しては、遺伝子水平伝達は、偶然の要素によって起こる。特定の機能のみ遺伝子水平伝達が起こり、その機能の遺伝子群がそっくり入れ替わるとは考えにくいと結論づけられる。

## 【 0 0 5 3 】

また、古細菌が真正細菌に進入し、共生した結果ではないかとする説に対しては、ある特定の機能に関する遺伝子群が置き換わったということは、その遺伝子群だけが置き換わりやすい特殊な環境であったことを示す。細胞内共生による核形成のプロセスはこの条件を満たし、またミトコンドリアや葉緑体の進化的形成が細胞内共生によっていることが広く認知されていることからもっともらしいといえると結論付けられる。従って、古細菌が真正細菌に進入し、共生した結果、真核生物が生じたと考えられる。

## 【 0 0 5 4 】

具体的には、真正細菌に細胞内共生した古細菌は真正細菌が合成する代謝産物を利用し続けるにつれ、それらの合成に必要な遺伝子群を失った。一方真正細菌は細胞周期などを古細菌に支配され、最終的には遺伝子複製からタンパク質合成までのプロセスも古細菌に依存し、それ以外の遺伝子は徐々に移行したと推測できる。このことを模式的に示したのが図 9 である。図 9 に示すように、細胞周期調節、核形成、DNA 複製、RNA 転写、小胞体形成、翻訳、リボゾーム形成などに関連する核関連の遺伝子 9 1 は、古細菌遺伝子が起源であると考えられ、細胞への物質輸送、代謝、ストレス応答、解毒、イオンホメオスタシスなどに関連する細胞質 9 2 は、真正細菌遺伝子を起源とする遺伝子群に関連する。また、ミトコンドリア形成、ミトコンドリア輸送などに関連するミトコンドリア 9 3 は、ミトコンドリアの共生に由来していると考えられる。

## 【 0 0 5 5 】

このように、遺伝子群をドメインレベルにまで細分化することによる相同性解析の検出能の上昇と、祖先型遺伝子をタンパク質として再現し、その活性を確認することで、遺伝子の由来に関する推測の信頼性が格段に向上する。

#### 【 0 0 5 6 】

以上説明したように、本実施形態によれば、全ゲノム配列が明らかになっている酵母とバクテリアの O R F データを用い、真核生物、真正細菌、古細菌の 3 つの生物群における進化的な関係を明らかにすることができる。また、ヒトやその他の生物の O R F データを用いてその相同性を解析することにより、あらゆる生物同士の相同性をしきい値の設定条件に依存せずに安定して極めて簡便に判定することができる。

#### 【 0 0 5 7 】

本発明は上記実施形態に限定されるものではない。B L A S T 法を用いて相同値ヒット数を算出したが、他の相同値算出法を適用することも可能であることはもちろんである。また、T 検定では、有意水準を 5 % に設定したが、これに限定されるものではなく異なる値の有意水準に基づいて検定してもよい。また、T 検定以外の検定手法を用いて検定してもよい。

#### 【 0 0 5 8 】

また、例えば、非常に多数の遺伝子集団を用いて、ある目的に従って比較したり、特定の遺伝子を選別したり、あるいは選別された遺伝子を具体的に特定することに本発明を適用することができる。すなわち、対象とする遺伝子グループが複数ある場合には、それぞれのしきい値（E 値）で計算することにより、非常に多数の遺伝子を有する複数のグループ間におけるそれぞれの遺伝子の、ある特徴を有する遺伝子群との相同性類似性を比較すること、さらにそこで対象となった遺伝子を抽出、特定することを非常に簡便に行うことができる。さらに、対象とした遺伝子群の中に、そのグループの中では特異な特徴を有する遺伝子を特定することもできる。例えば、生物の進化過程においては親から子への伝達ではない、いわゆる水平伝達で導入された遺伝子の抽出にも適用可能である。

#### 【 0 0 5 9 】

特定された遺伝子を解析した生物の中から取り出し、D N A 塩基配列を知るこ

と、あるいは類似の遺伝子を検索し、類似の遺伝子の情報に基づきそれらの遺伝子の機能を推測することが可能となる。また、ある目的に従った人材の選別などにも応用できる。すなわち、非常に多数のデータを含むグループ（データ群）の中から、ある一定の条件を満たす、又はその条件の度合いを変化させつつデータを選別、特定できる。グループ間で相互に相同性を解析することにより、お互いの条件を満たすペアを選別することもできる。

#### 【0 0 6 0】

また、現在、ヒトゲノム解明以後の大きな流れとして、病気の原因となる遺伝子の選別特定が重要となっている。候補となっている遺伝子を実際のヒトDNA試料からPCR遺伝子増幅法により、DNA断片として作成し、大腸菌等の発現系でその遺伝子が合成するタンパク質を実際に作らせ、その活性をみることや或いはそのタンパク質に対する抗体を作成することが行われる。抗体は実際にヒトの病理切片などで試すことで病気の検出方法等に有用である。また、病気の原因となる遺伝子の特定は患者の組織の病域の特定、或いは健常人の遺伝子を解析することでその病気にかかりやすいヒトを選別することができる。そうすることで、患者に対する治療の方法、例えば、その保因者とそうでない人に対しては、例えばポリープ切除の緊急性の判断に応用できる。健常人に関してはその病気に関する診断検査のそれぞれの人の遺伝子情報に基づき使い分けることが可能である。

#### 【0 0 6 1】

また、相同性比較の対象とするデータ群として、遺伝子以外のデータを用いることにより、人材情報データの解析、選別や、商品の類似性判定、気象情報解析などにも応用できる。すなわち、複数のデータからなるデータ群同士の類似性の度合いを解析するものであれば何でも適用可能である。

#### 【0 0 6 2】

##### （第2実施形態）

本実施形態は第1実施形態の変形例に関わる。本実施形態は、第1実施形態のデータの相同性解析を取引の成立の仲介サービスに適用する形態に関わる。本実施形態では、人材斡旋を例に説明する。



**【0063】**

図10は本実施形態の取引成立システムの全体構成を示す図である。図10に示すように、本実施形態の人材斡旋を行う人材斡旋業者が運営するサーバ101と、この人材斡旋サービスにより希望する人材の獲得を希望する求人者が利用する求人者端末102と、この人材斡旋サービスにより希望する企業への就職を希望する求職者が利用する求職者端末103がインターネット104に接続されている。

**【0064】**

インターネット104はデータの送受信により情報のやりとりを行ういかなるネットワークに置き換えて実施してもよい。また、その接続形態は有線、無線の別を問わない。

**【0065】**

図11はサーバ1の詳細な構成を示す図である。図11に示すように、プロセッサ111と、このプロセッサ111に接続され、ネットワーク4との間の情報の送受信を制御するインタフェース112と、プロセッサ111に接続され、求人者から得られた求人データを格納する求人データ記憶手段113と、求職者から得られた求職者データを格納する求職データ記憶手段114から構成される。プロセッサ111は、相同性解析手段111aと検定手段111bを有する。相同性解析手段111aは、複数のデータからなるあるデータ群と他のデータ群の相同性を相同値として算出する相同値算出手段を備える。また、検定手段111bは第1実施形態で示したT検定手段12と同様の構成であり、相同性決定値算出手段、相同性可否決定値算出手段及び判定結果導出手段を備える。

**【0066】**

次に、本実施形態の人材斡旋方法を説明する。

**【0067】**

まず、サーバ1は、人材斡旋WEBページをインターネット104を介して求人者や求職者などが閲覧可能な状態にサーバ1に設定する。より具体的には、人材斡旋WEBページファイルをサーバ1の図示しない記憶手段にインターネット104を介してデータ取得可能に格納しておく。なお、この人材斡旋WEBペー

ジファイルは求人者用WEBページファイルと求職者用WEBページファイルがあり、求人者からの接続要求に対しては求人者用WEBページファイルが、求職者からの接続要求に対しては求職者用WEBページファイルが送信されるようになっている。

#### 【0068】

求人者が求人者端末102からサーバ1の人材斡旋WEBページへの接続要求を行うと、サーバ1は人材斡旋WEBページファイルを求人者端末102に送信する。求人者端末102の図示しない表示手段には、例えば図12(a)に示すような求人者用WEBページが表示される。求人者は、この図12(a)に示す画面に従って、求人者端末102を用いて求人データを入力する。求人データは、例えば年齢、性別、職種、業種、地域、通勤時間、希望する賃金、一年間の休暇日数、希望する経験や能力、例えば資格試験などの点数、勤務の形態や年数などを入力する。入力した求人データは求人者端末102に送信される。サーバ101は受信した求人データを求人データ記憶手段113に登録する。なお、求人データを構成する各データ項目は、求人データ記憶手段113に登録される際に数値データに変換されるのが望ましい。また、各データ項目は、例えば月収であれば25万円というように一義的な値として示してもよいし、25万円～30万円というように、所定の幅をもった値としてもよい。

#### 【0069】

一方、求職者が求職者端末103からサーバ1の人材斡旋WEBページへの接続要求を行うと、サーバ1は人材斡旋WEBページファイルを求職者端末103に送信する。求職者端末103の表示手段には、例えば図12(b)に示すような求職者用WEBページが表示される。求職者は、この図12(b)に示す画面に従って、求職者端末103を用いて求職データを入力する。求職データを構成する各データは求人データと共通する。入力した求職データは求職者端末103に送信される。サーバ101は受信した求職データを求職データ記憶手段114に格納する。なお、求職データを構成する各データ項目は、休職データ記憶手段114に登録される際に数値データに変換されるのが望ましい。また、各データ項目は、例えば月収であれば25万円というように一義的な値として示してもよ

いし、25万円～30万円というように、所定の幅をもった値としてもよい。

#### 【0070】

以上のような工程を経て複数の求人データと複数の求職データがサーバ1がそれぞれ求人データ記憶手段113と求職データ記憶手段114に登録されると、サーバ1の相同性解析手段111aは解析対象として例えばある求人者の求人データを選択し、各求職データとの相同値ヒット数を算出する。この相同値ヒット数の算出は、第1実施形態のBLAST解析手段11に処理と同様である。第1実施形態における酵母ORFを本実施形態のある求人者の求人データを構成する各データ項目に置換することにより、求人データの各求職データとの相同値ヒット数が算出される。

#### 【0071】

但し、本実施形態の人材斡旋の場合、各データ項目は第1実施形態のORFのように、断片的な情報として順序づけられていない情報ではない。従って、例えば月収であれば月収同士を、通勤時間であれば通勤時間同士についてヒットするか否かを判定すればよい。図13は、相同性の有無の判定の概念を示す図である。図13に示すように、図5とは異なり、異なる種別のデータ項目同士のヒット数の判定は行う必要が無い。

#### 【0072】

このような相同値ヒット数の算出を第1実施形態と同様に複数のしきい値を用いて、かつすべての求職データについて行い、得られた相同値ヒット数をその解析対象とした求人データに関連づけて求人データ記憶手段113に格納する。

#### 【0073】

以上のようにして得られた相同値ヒット数としきい値との関係は、第1実施形態の図6と同様に得られる。ある求人者に対する各求職者との関係が折れ線グラフで示される。横軸は、求人データの各データ項目の条件の厳しさを示しており、縦軸はどの求職者がその条件を満たすかを示している。この相同値ヒット数としきい値との関係を示すデータを解析対象とした求人者端末102に送信し、求人者端末102側で表示手段に表示させることにより、求人者は必要とする人材を選別することができる。

**【 0 0 7 4 】**

なお、上記では一旦求人データを登録した場合に、その条件を変動させずに相同性判定を行う例を示したが、例えば月収など、各データ項目の条件を変動させると、それに応じて各求職者のヒット数に変動する。このヒット数の変動に伴い、求職者それぞれの折れ線グラフが変動する。このように、データ項目の条件を変動させることにより変動する候補を選別するのが容易となる。

**【 0 0 7 5 】**

また、以上のようなある求人者の求人データに対する相同性判定をある求職者の求職データに対して行うことにより、その求職者にとって条件に見合った求人を選別することができる。この場合も、求人データに対する相同性判定と同様に、図 6 に示すような相同値ヒット数としきい値との関係曲線を得ることができる。このような相同値ヒット数としきい値との関係を示すデータを解析対象とした求職者端末 1 0 3 に送信し、求職者端末 1 0 3 側で表示手段に表示させることにより、求職者は必要とする人材を選別することができる。

**【 0 0 7 6 】**

以上は相同値ヒット数としきい値との関係に基づいて人材の選別を行う場合を示したが、図 6 に示すような関係曲線のみでは選別が困難な場合も多い。そこで、プロセッサ 1 の検定手段 1 1 1 b は、第 1 実施形態の T 検定処理と同様の検定処理を実行する。検定処理は、第 1 実施形態の T 検定手段 1 2 と同様の相同性決定値の算出、相同性可否決定値の算出及び判定結果導出の 3 つの処理に加えて、順序付け処理からなる。相同性決定値の算出、相同性可否決定値の算出及び判定結果導出については第 1 実施形態と同様であるので詳細な説明は省略する。

**【 0 0 7 7 】**

判定結果導出処理では、いずれのデータ群に相同性を有するかを判定することができる。これにより、ある条件以上に合った人材を選別することは可能である。

**【 0 0 7 8 】**

一方、求人者は、どの人材が最も条件に見合っているか、また求職者は、どの人材が最も条件に見合っているか、さらにはどのような順序で人材が自身の提示

した条件に合致してくるかが分かると人材登用や就職活動に有益である。

#### 【 0 0 7 9 】

そこで、判定結果導出処理で得られたデータに基づき、人材の条件に対する順序付け処理を実行する。例えばある求人者に対する各求職者の判定結果に基づき順序付け処理を行う場合、判定結果導出処理により、図 1 4 に示すような判定結果が得られる。そこで、この判定結果に基づき、相同性の高い順序に各求職者を順序付ける。順序づけは、例えばある求職者をある順位と仮定し、その仮定に対して判定結果を組み合わせしていくことで、簡単に得られる。図 1 4 の例で求職者 a を 1 位とした場合、レコード 1 の情報によれば b の方が順序が上であるため、求職者 b を 1 位として求職者 a を 2 位とする。レコード 2 の情報によれば c の方が a より順序が上であるため、求職者 c を 1 位あるいは 2 位とし、求職者 b を 2 位あるいは 1 位とし、求職者 a を 3 位とする…というように、各判定結果に基づき、順序付けを行うことができる。得られた順序付け処理結果は図 1 4 (b) のようになる。

#### 【 0 0 8 0 】

この順序付け処理は、ある求人者の求人データに基づく求職者の順序付けのみならず、ある求職者の求職データに基づく求人者の順位付けにも適用できることはもちろんである。

#### 【 0 0 8 1 】

このような順位付け処理結果を求人者端末 1 0 2 や求職者端末 1 0 3 に、相同値ヒット数としきい値との関係曲線を示すデータとともに送信することにより、人材の選別がさらに容易になる。もちろん、求人データや求職データの各データ項目の条件を変動させる場合に、条件の変動に対応して順位付けを行うことにより、条件を変動させた場合に人材が提示した条件に対してどのように変動させるかを把握することができ、人材選別がさらに容易になる。

#### 【 0 0 8 2 】

なお、本実施形態では求人者端末 1 0 2 と求職者端末 1 0 3 からインターネット 1 0 4 を介してデータを収集し、そのデータに基づいて人材斡旋処理を行う例として示したが、これに限定されるものではない。求人者と求職者からネットワ

ークを介さずに電話、F A X、紙媒体などにより情報を受け取り、その情報をサーバ 1 に入力することにより、上記と同様の人材斡旋処理が可能となる。

#### 【 0 0 8 3 】

また、本実施形態は人材斡旋処理という目的に限定して取引成立システムを説明したが、人材斡旋のために必要な求人データや求職データを他のデータに置換することにより、人材以外の他の商品やサービスなどの取引を行うシステムに容易に応用可能である。

#### 【 0 0 8 4 】

例えば、商品の取引システムに応用する場合、商品を購入あるいは販売するための条件（例えば商品の種類、商品の機能、商品の価格、商品が得られるまでの時間など）と上記求人データや求職データを置換すればよい。また、結婚相談サービスを提供する場合には、男性と女性の結婚相手に求める各種条件（例えば年齢、年収、身長、体重、性格、住所、家族構成など）と上記求人データや求職データを置換すればよい。

#### 【 0 0 8 5 】

いずれの取引システムに本実施形態を適用する場合でも、例えば数値化が困難な情報については、概念や条件として近いデータ同士に差分の少ない数値をそれぞれに割り当てることにより、相同性判定の精度を向上させることができる。例えば、人材斡旋における職種は数値化が困難である。職種として例えば製造業、金融業、保険業がある場合に、金融業と保険業は比較的隣接する概念であり、いずれを任意に選択しても良いと思う求職者が多いのに対して、製造業はこれら金融業と保険業とは比較的隣接しない概念である。このような場合、金融業と保険業には数値“1”と“2”を割り当て、製造業には“1”や“2”とは大きく離れた数値“20”を割り当てればよい。これにより、数値の値の大きさにより相同性を判定することができる。また、このような数値の割り当て法以外にも、データを選択式にして求人者や求職者に複数のデータを選択させ（例えば金融業と保険業双方）、選択されたものが一致するか否かによりヒット数を算出するなどにより、相同性の判定の精度を高めることができる。

#### 【 0 0 8 6 】

また、取引の成立のみならず、男性及び女性の結婚希望者同士のマッチングなど、2つの異なる集団間の相互の希望のマッチングを行うあらゆるシステムに本実施形態を応用可能である。もちろん、一方の集団は単一でもよいので、オークションなどにも応用可能である。

#### 【0087】

本発明は、さらに別の技術にも適用可能である。例えば、衛星写真などの画像データを複数の時刻について取得し、それぞれの相同性を上記実施形態と同様に判定することにより、変化した部分を容易に特定する技術に適用可能である。また、例えば火災を撮像して比較すれば、火災検出が可能となり、火災防止に応用できる。また、航空機の管制塔などに於けるレーダー上に刻々と変わる飛行機などの情報を相互に比較し、新たな変化などを瞬時に選別することもできるし、ミサイルなどの発射による地点のわずかな変化を瞬時に検出することにも応用できる。また、画像として天気図を取得した場合には、過去の天気図と比較することにより天気の予測が可能となる。また、天気図のみならず、その画像として取り込まれた各地点の気温、気圧、気象条件などを併せて取得し、それらについても過去のデータと対比して相同性判定を行うことにより、より正確な天気の予測が可能となる。

#### 【0088】

また、多数の指紋や似顔絵などの画像をそれぞれの部分の特徴などを分解し、その構成をパターン化したり、最も類似するものを選別、特定する。その際、各要素に対して許容度を変動させながら選別される事例を取り出すことにより、犯人の特定や人類学的な相互比較などに応用できる。

#### 【0089】

また、画像データをデジタル信号としてそれぞれ分割し、分割したデジタル信号同士に対して相同性判定を行うこともできるし、この相同性判定の対象とする信号を他の信号に置換することにより、2つの信号を比較し、異なる信号となっている部分を容易に特定することが可能である。例えば、音声信号などを所定の周波数で分割して比較することにより、音楽の類似性を判定することが可能である。また、例えば小説などの文字情報を所定の長さで分割して比較することによ

り、小説などにおける言い回しや使用する単語の類似性を判定することができる。これにより、対比する2つの創作の類似性や盗作の可能性などに応用することができる。もちろん、このような創作性の判定のみならず、言語学の分野では、子供から大人へ、あるいは地域ごとの言語の相違点などを相互に比較し、その特徴を抽出することにより、共通な点や異なる点を具体的に要素として抽出することが可能となる。

#### 【0090】

#### 【発明の効果】

以上詳述したように本発明によれば、多数のデータ同士を定量的にかつ正確に比較することが可能となる。

#### 【図面の簡単な説明】

#### 【図1】

本発明の第1実施形態に係る遺伝子解析システムの全体構成を示す図。

#### 【図2】

同実施形態に係る遺伝子解析に必要な遺伝子データを取得するプロセスの一例を示す図。

#### 【図3】

同実施形態に係るアミノ酸配列記憶手段に格納された遺伝子データ群の概念図。

#### 【図4】

同実施形態に係るBLAST法の概念図。

#### 【図5】

同実施形態に係るBLAST法による相同性の有無の判定の概念を示す図。

#### 【図6】

同実施形態に係る相同値ヒット数としきい値Eとの関係を示す図。

#### 【図7】

同実施形態に係る相同性決定値 $Z_i(1)$ を示す図。

#### 【図8】

同実施形態に係るT検定における判定結果導出に用いられる判定テーブルの一



例を示す図。

【図 9】

同実施形態に係る T 検定に於ける判定結果により考察した酵母遺伝子の由来を模式的に示す図。

【図 1 0】

本発明の第 2 実施形態の取引成立システムの全体構成を示す図

【図 1 1】

同実施形態に係るサーバの詳細な構成を示す図

【図 1 2】

同実施形態に係る人材斡旋 W E B ページの表示例を示す図。

【図 1 3】

同実施形態に係る相同性の有無の判定の概念を示す図。

【図 1 4】

同実施形態に係る順序づけ処理を説明するための図。

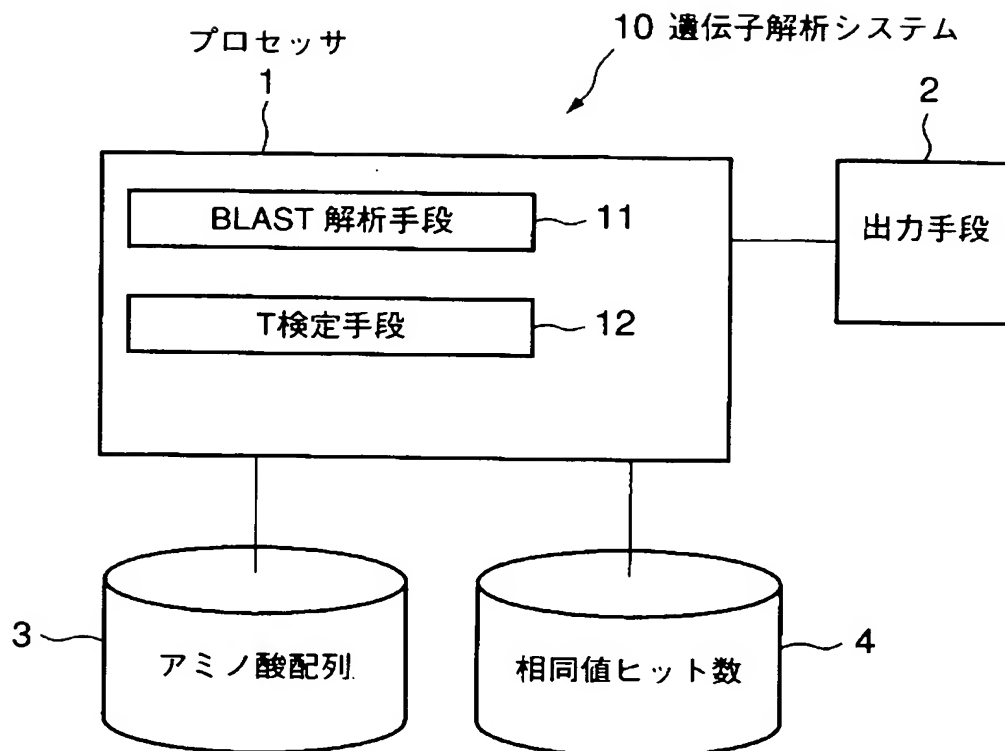
【符号の説明】

- 1 … プロセッサ
- 2 … 出力手段
- 3 … アミノ酸配列記憶手段
- 4 … 相同値記憶手段
- 1 1 … B L A S T 解析手段
- 1 2 … T 検定手段
- 1 0 1 … サーバ
- 1 0 2 … 求人者端末
- 1 0 3 … 求職者端末
- 1 0 4 … インターネット

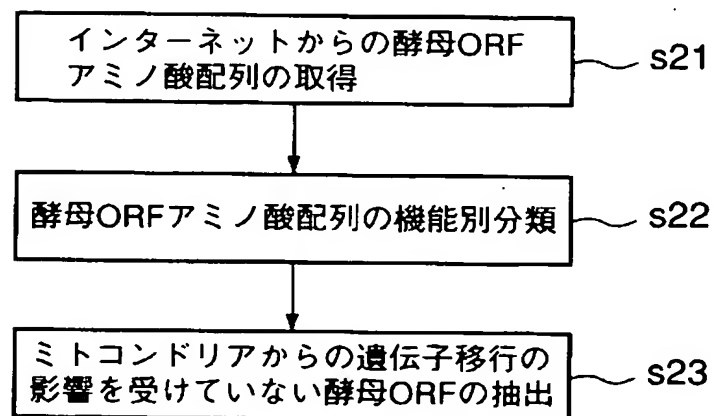
【書類名】

図面

【図 1】



【図 2】



【図 3】

酵母遺伝子データ					
ORF1	ORF2	ORF3	ORF4	ORF5	ORFp
アミノ酸a,b,c	アミノ酸a,b,d	アミノ酸c,d,e	アミノ酸f,g,h	アミノ酸z,r,t,g,h	アミノ酸e,f,g,b,x

古細菌A遺伝子データ					
ORF1	ORF2	ORF3	ORF4	ORF5	ORFn1
アミノ酸d,y,e	アミノ酸w,x,b	アミノ酸x,d,b	アミノ酸w,x,d	アミノ酸s,d,h	アミノ酸x,f,y

古細菌B遺伝子データ					
ORF1	ORF2	ORF3	ORF4	ORF5	ORFn2
アミノ酸d,e,s	アミノ酸z,d,s	アミノ酸x,b,j	アミノ酸b,y,i	アミノ酸p,b,c	アミノ酸t,y,x

.

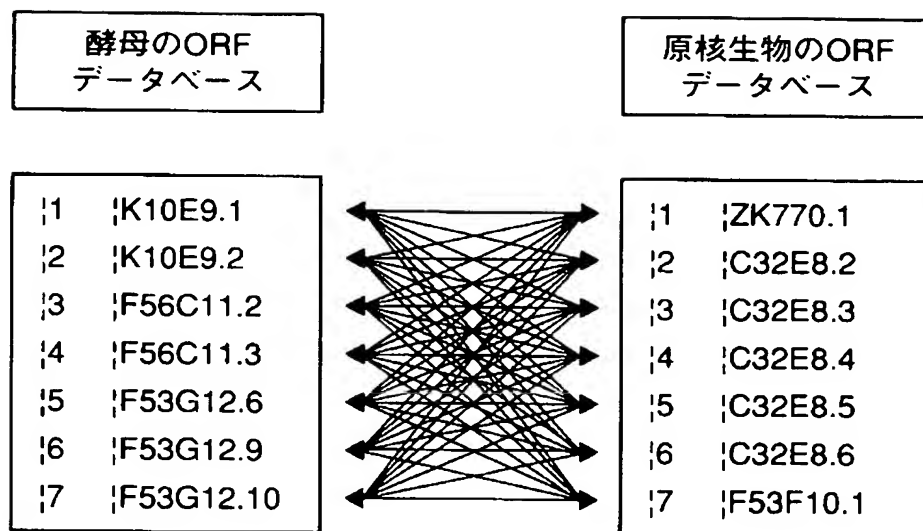
.

.

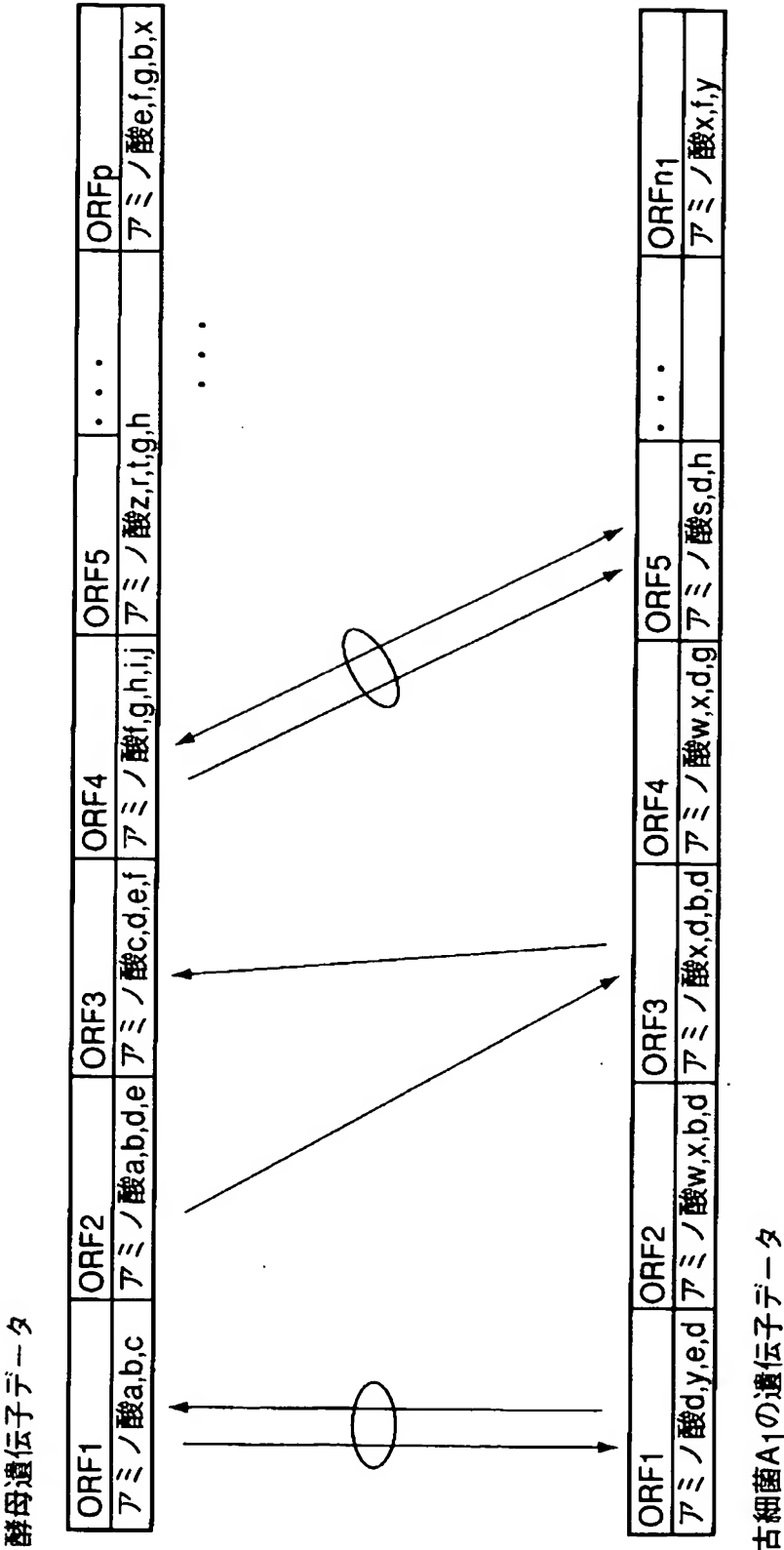
真性細菌C遺伝子データ					
ORF1	ORF2	ORF3	ORF4	ORF5	ORFn3
アミノ酸t,b,g	アミノ酸x,f	アミノ酸q,c	アミノ酸n,m	アミノ酸x,l	アミノ酸w,r

真性細菌D遺伝子データ					
ORF1	ORF2	ORF3	ORF4	ORF5	ORFn4
アミノ酸y,c	アミノ酸w,e,r	アミノ酸q,x,t	アミノ酸p,b	アミノ酸q,n,c	アミノ酸o,v,u

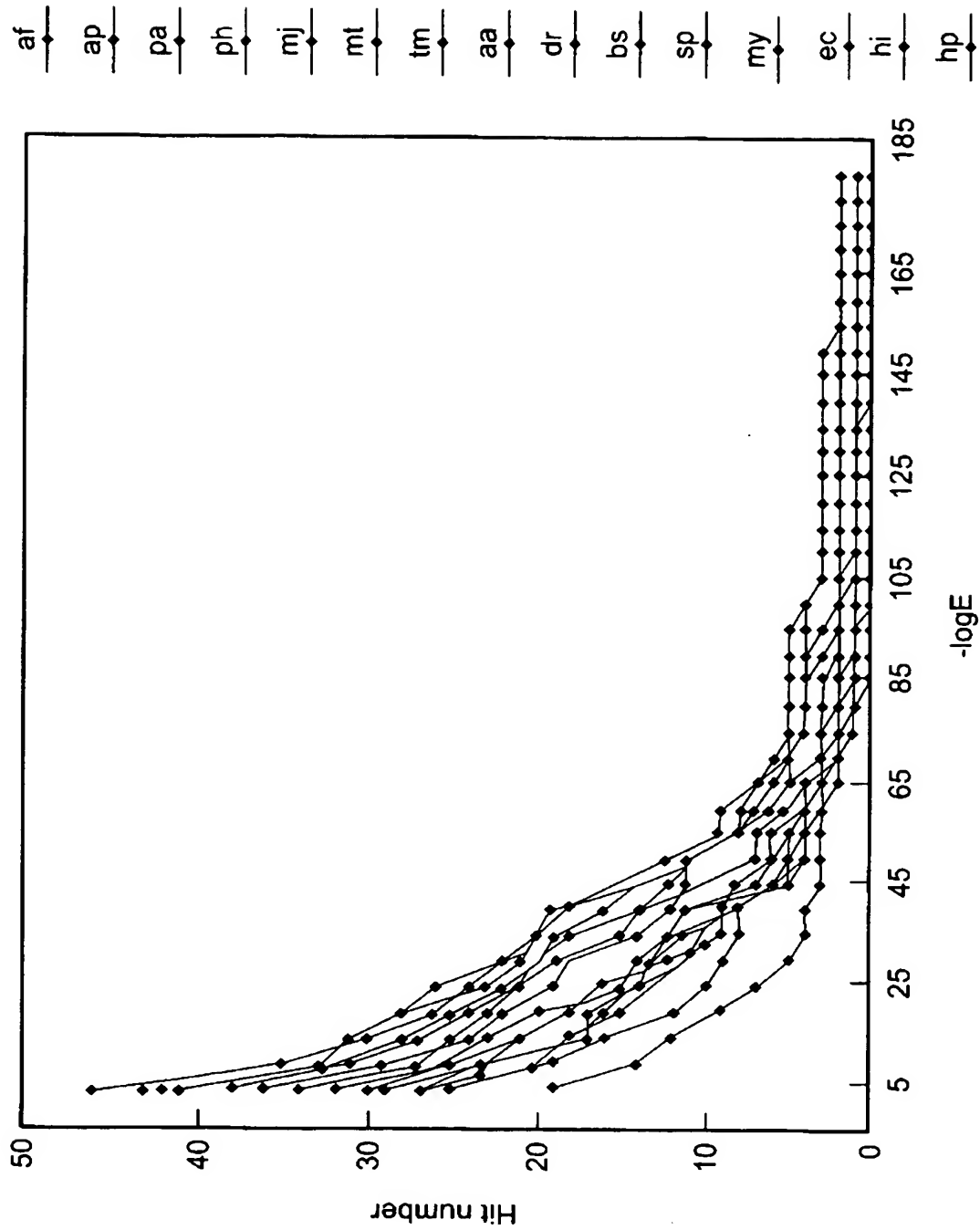
【図 4】



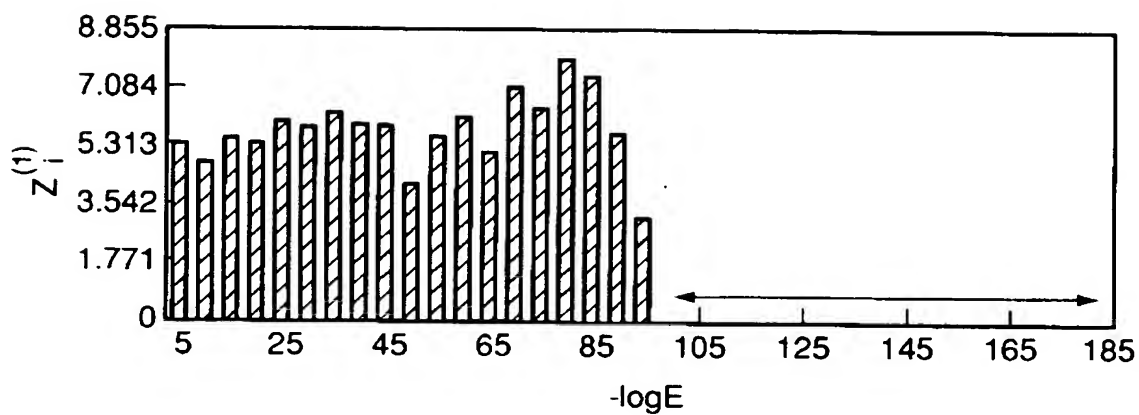
【図 5】



【図 6】



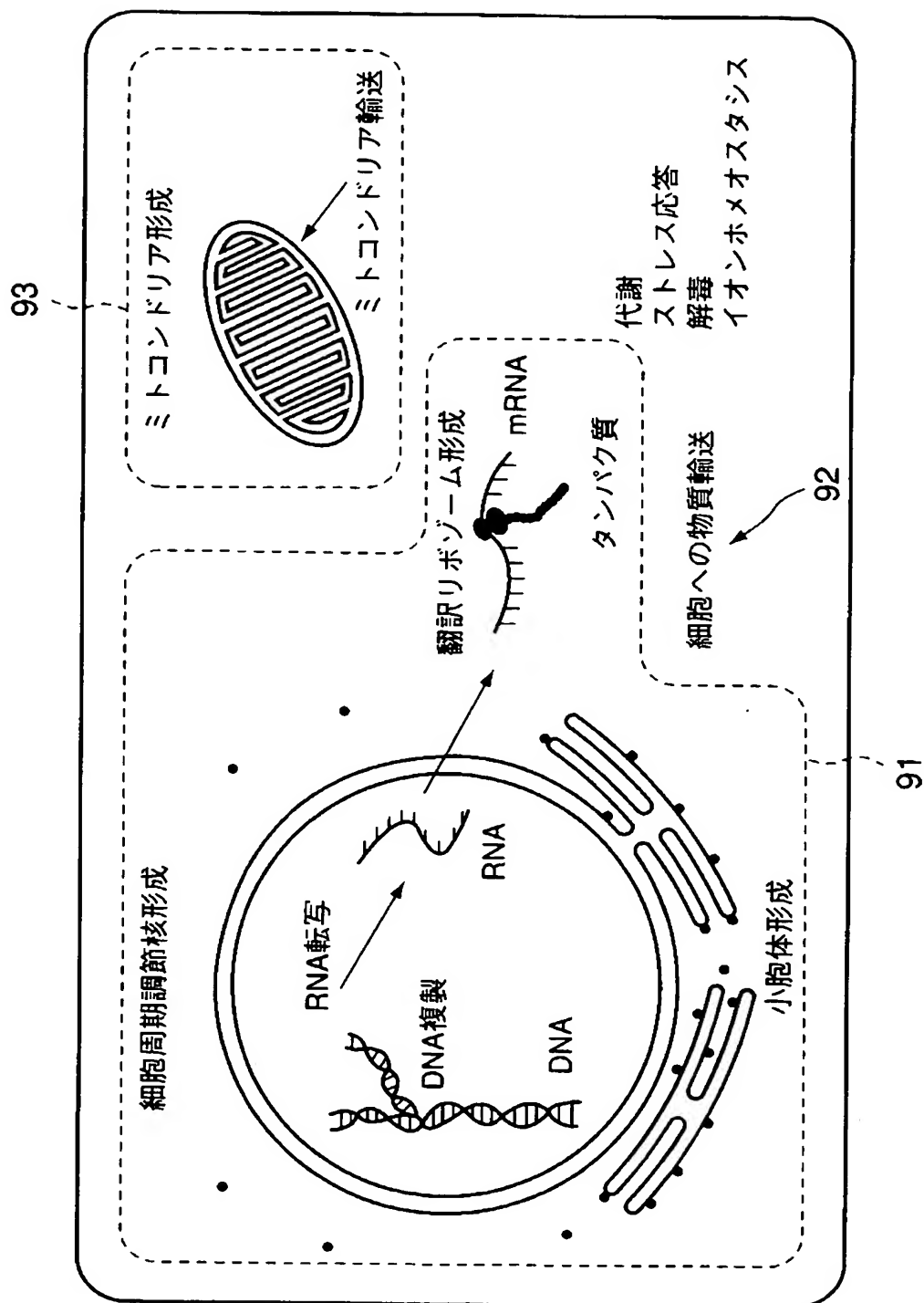
【図 7】



【図 8】

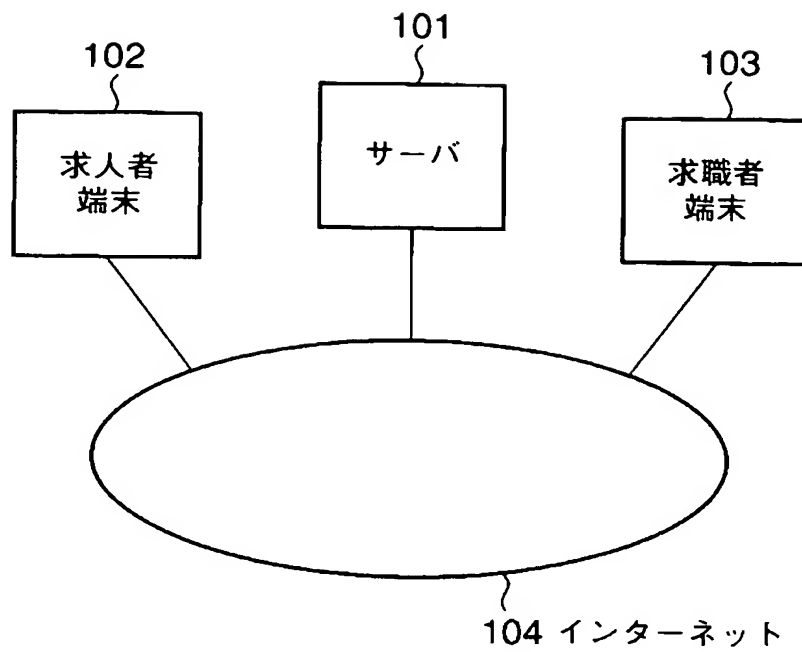
$Z^{(2)}/\ln-1$ \ $\overline{Z^{(1)}}$	+	-
1以上	古細菌	真正細菌
1未満	判定不可	判定不可

【図 9】

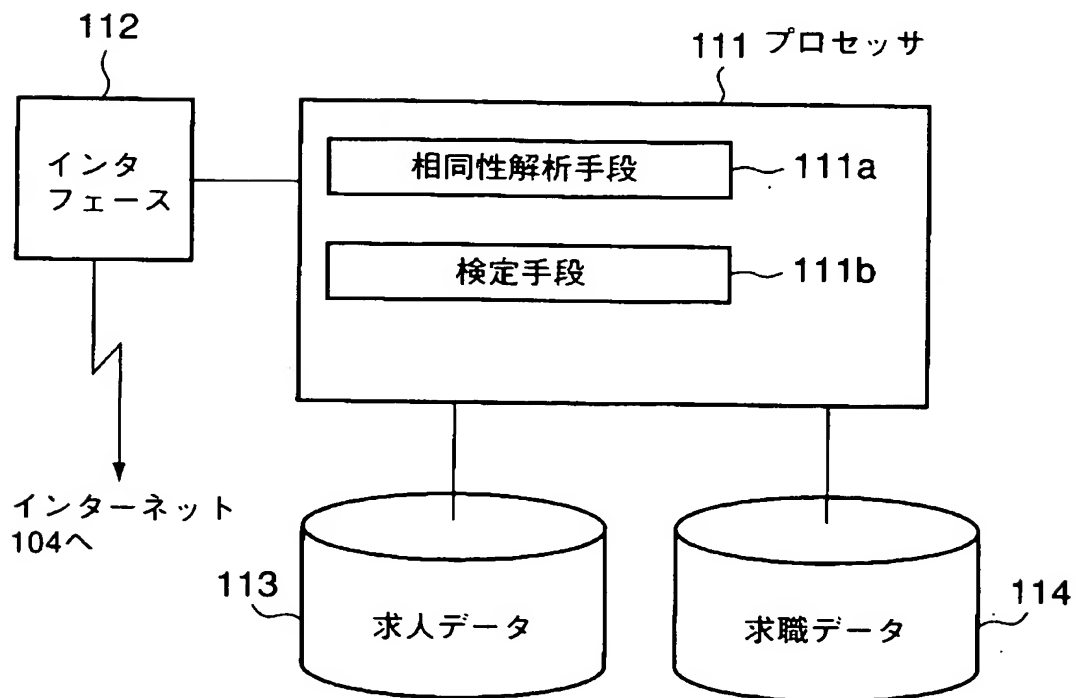




【図10】



【図11】



【図 12】

求人者の方は、必要とされる人材の条件を以下に入力して下さい。

年齢	歳～ 歳
性別	男性 <input type="radio"/> 女性 <input type="radio"/>
職種	
業種	
地域	関東 <input type="radio"/> 近畿 <input type="radio"/> 四国 <input type="radio"/>
通勤時間	時間～ 時間
希望する賃金	円～ 円
一年間の休暇日数	日～ 日
希望する経験、能力	
資格試験の点数	点～ 点
勤務形態、年数	

入力が終わりましたら  右のボタンをクリックして下さい。

(a)

求職者の方は、必要とされる人材の条件を以下に入力して下さい。

年齢	歳～ 歳
性別	男性 <input type="radio"/> 女性 <input type="radio"/>
職種	
業種	
地域	関東 <input type="radio"/> 近畿 <input type="radio"/> 四国 <input type="radio"/>
通勤時間	時間～ 時間
希望する賃金	円～ 円
一年間の休暇日数	日～ 日
希望する経験、能力	
資格試験の点数	点～ 点
勤務形態、年数	

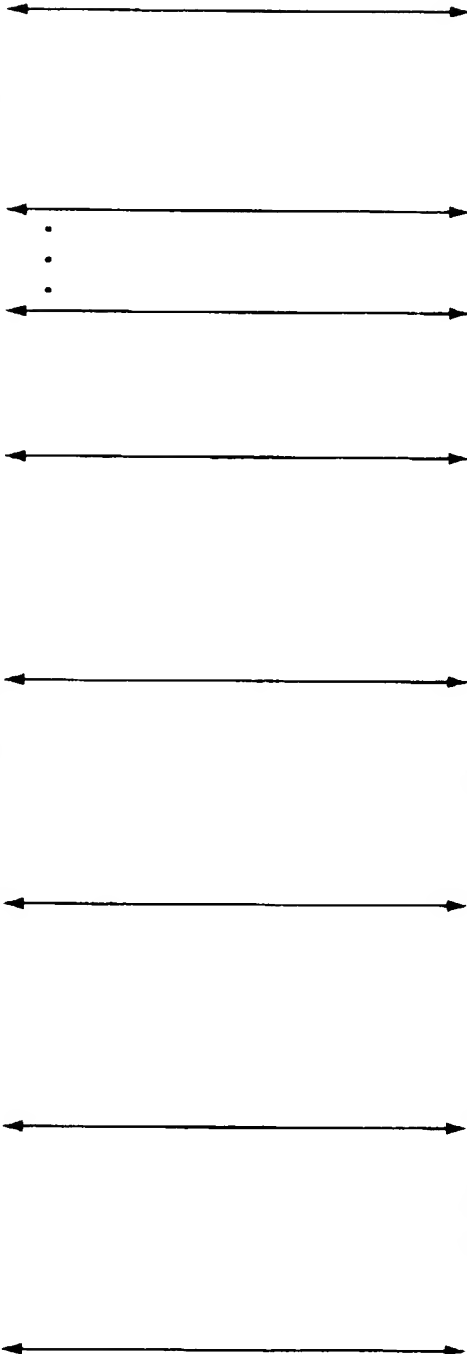
入力が終わりましたら  右のボタンをクリックして下さい。

(b)

【図 1 3】

求人データ

年齢	性別	職種	業種	通勤時間	...	資格試験の点数
23～30	1 (男)	3 (金融)	4 (経理)	1時間～1時間半	...	600～



年齢	性別	職種	業種	通勤時間	...	資格試験の点数
31	1 (男)	4 (保険)	5 (財務)	30分～1時間	...	750

求職データ

【図 14】

レコード数	比較データ	判定結果
レコード1	a,b	b
レコード2	a,c	c
レコード3	a,d	d
レコード4	b,c	b
レコード5	b,d	b
レコード6	c,d	d

(a)

順位付け結果
b
d
c
a

(b)

【書類名】 要約書

【要約】

【課題】 多数のデータ群同士を定量的にかつ正確に比較する。

【解決手段】 B L A S T 解析手段 1 1 により、解析対象データ群と第 1 のデータ群のそれぞれのデータ群に含まれるデータの相同性を計算し、その値を用いて相同性有りとなるデータの数（オルソログス遺伝子数）を第 1 の相同値  $x$  として算出し、しきい値  $E$  を  $n_A$  個設定して各しきい値  $E_i$  毎に第 1 の相同値  $x_i$  を算出し、解析対象データ群と第 2 のデータ群のそれぞれのデータ群に含まれるデータの相同性を計算し、相同性有りとなるデータの数（オルソログス遺伝子数）を第 2 の相同値  $y$  として算出し、しきい値  $E$  を  $n_B$  個設定して各しきい値  $E_i$  毎に第 2 の相同値  $y_i$  を算出し、T 検定手段 1 2 は、第 1 の相同値  $x_i$  及び第 2 の相同値  $y_i$  並びにしきい値の数  $n$  との関係に基づいて、解析対象データ群が第 1 のデータ群又は第 2 のデータ群のいずれに類似するかを判定する。

【選択図】 図 1

## 認定・付加情報

特許出願の番号	特願 2 0 0 1 - 2 3 1 8 1 0
受付番号	5 0 1 0 1 1 2 5 3 8 8
書類名	特許願
担当官	佐藤 浩聡 7 6 6 4
作成日	平成 1 4 年 1 月 1 5 日

## &lt; 認定情報・付加情報 &gt;

## 【特許出願人】

【識別番号】	396020800
【住所又は居所】	埼玉県川口市本町 4 丁目 1 番 8 号
【氏名又は名称】	科学技術振興事業団

## 【代理人】

申請人

【識別番号】	100058479
【住所又は居所】	東京都千代田区霞が関 3 丁目 7 番 2 号 鈴榮内外 国特許法律事務所内
【氏名又は名称】	鈴江 武彦

## 【選任した代理人】

【識別番号】	100092196
【住所又は居所】	東京都千代田区霞が関 3 丁目 7 番 2 号 鈴榮内外 国特許法律事務所内
【氏名又は名称】	橋本 良郎

## 【選任した代理人】

【識別番号】	100091351
【住所又は居所】	東京都千代田区霞が関 3 丁目 7 番 2 号 鈴榮内外 国特許法律事務所内
【氏名又は名称】	河野 哲

## 【選任した代理人】

【識別番号】	100088683
【住所又は居所】	東京都千代田区霞が関 3 丁目 7 番 2 号 鈴榮内外 国特許法律事務所内
【氏名又は名称】	中村 誠

次頁無

特願 2001-231810

出 願 人 履 歴 情 報

識別番号

[396020800]

1. 変更年月日

1998年 2月24日

[変更理由]

名称変更

住 所

埼玉県川口市本町4丁目1番8号

氏 名

科学技術振興事業団